

Metadata and EPOS-DCAT-AP

- Author: Tim Sonnemann (tim@vedur.is)
- Created: 2018-02-13

Purpose of this document

- Explain EPOS-DCAT-AP.

Summary

- **EPOS-DCAT-AP** is an extension of **DCAT-AP** is an extension of **DCAT**.
- These are all **metadata vocabularies** to describe (web)resources/things/contacts.
- **DCAT** is recommended by the **W3C** for global use to enable linking basic types of datasets.
- **DCAT-AP** was developed for data portals in Europe to improve discoverability.
- **EPOS-DCAT-AP** was developed as extension for Solid Earth data in the **EPOS** project.

About Metadata

Mostly taken from Tim Berners-Lee's article about [Metadata Architecture](#):

- Information about information is generally known as **Metadata**.
- Definition: Metadata is **machine understandable** information about web resources or other things.
- Metadata is data.
- Metadata about one document can occur within the document, or within a separate document, or it may be transferred accompanying the document.
- Metadata can describe metadata.
- Metadata consists of **assertions** about data (e.g. name="report_01.doc").
- Data have **attributes** (e.g. a file can have a *name*, which is one of its attributes).
- Attributes specific to one type of data can be defined in a dedicated **name space**.
- Vocabulary name spaces can be defined/stored at a specific URL.

DCAT Summary

- **DCAT** is a **W3C vocabulary recommendation for metadata** to describe datasets in data catalogs across different websites and allow their general interoperability and discoverability on the web.
 - **W3C (World Wide Web Consortium)** is the main international standards organization for the World Wide Web.
 - A **DCAT profile** is a specification for data catalogs that **extends DCAT**.
- DCAT is an **RDF** vocabulary suited to representing government data catalogs.
 - The **Resource Description Framework (RDF)** is a language for representing information about resources in the World Wide Web.
- It's currently one of the most widely used Semantic Web vocabularies for describing datasets and data catalogues.
- What is defined in DCAT?
 - DCAT is basically a database schema with attributes defined by various name spaces.
 - There are 7 linked classes (tables) total.
 - Each class (table) and attribute (field) has naming pattern `<namespace>:<property>`
 - DCAT uses already existing name spaces for most properties:
 - **dct** = **Dublin Core** terms
 - **dctype** = Dublin Core object types
 - **foaf** = "Friend of a Friend" dictionary for human relations
 - **rdf** = RDF vocabulary terms
 - **rdfs** = RDF Schema vocabulary
 - **skos** = Simple Knowledge Organization System
 - **vcard** = contact info (virtual business card)
 - **xsd** = XML schema definition
 - The **dcat** vocabulary specific to DCAT has 4 classes
 - **dcat:Catalog**
 - **dcat:CatalogRecord**
 - **dcat:Dataset** describes common info about a collection of resources
 - **dcat:Distribution** describes a specific resource (file)
 - Other classes used are:
 - **skos:ConceptScheme**
 - **skos:Concept**
 - **foaf:Agent** (can be **foaf:Person** or **foaf:Organization**)

DCAT-AP Summary

- **DCAT-AP: DCAT Application profile for data portals in Europe**
- DCAT-AP is a DCAT profile (i.e. extension) for European applications.
 - Metadata profile based on and compliant with DCAT of W3C.
- **Purpose:** Define a common interchange metadata format for data portals of the EU and of EU Member States.
- Developed in the framework of the EU Programme *Interoperability Solutions for European Public Administrations* (ISA)
- Specification: [DCAT-AP v1.1](#) (Oct 2015)
 - UML diagram on page 8
 - 1-page reference for all classes and properties on page 26
- **Differences** with DCAT:
 - DCAT-AP adds 15 classes and various properties to DCAT.
 - language, keywords, referenced standards, checksum, location, etc.
 - DCAT-AP also specifies the requirement type for every class/property:
 - mandatory, recommended, optional

EPOS-DCAT-AP Summary

- **EPOS-DCAT-AP** is a DCAT-AP extension for Solid Earth data.
- Developed in context of European Plate Observing System (**EPOS**) project
- **UML diagram** shows added classes and links around original DCAT-AP
 - WARNING: The diagram is not up to date and is missing some properties!
- **New classes** are used to describe:
 - Publication (abstract, issued, title, issn, ...)
 - Organisation (isPartOf, associatedProjects, scientific/legal/financial contact)
 - Person (language, affiliation, qualification)
 - Software (software application schema, model code)
 - Service (terms of use, service type, schema)
 - Webservice (EPOS domain, parameters, contact, created, ...)
 - Equipment (description, type, manufacturer, dynamic range, ...)
 - Facility (theme, type, description, foaf:page, ...)
 - Dataset (character encoding, subject, type, domain, subdomain, ...)
- **Mandatory properties** describe the minimum set of information needed by the Integrated Core Service (**ICS**) to automatically categorize, query and expose a dataset/service.
- The required information classes with mandatory/optional properties for **EPOS TCS-ICS integration** are listed and described in an Excel file: [EPOS_DCAT-AP_Vocabulary_and_Specification.xlsx](#)